

# A Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály kutatási célkitűzései a 2017. évre

## Nyelvtechnológiai Kutatócsoport

- 2017-ben lezárul a Finn-OTKA projekt, amelynek keretein belül megtörténik az automatikusan előállított protoszótárak kiértékelése. A létrehozott szótárak bizonyos morfológiai, etimológiai, szemantikai információkkal és többnyelvű fordítási megfelelőikkel kibővítve felkerülnek a Wiktionarybe. A feltöltendő Wiktionary-cikkek, amennyire lehet, automatikusan állnak elő, de az utolsó fázisban szükség lesz még kézi ellenőrzésre. Fontos az eredmények hazai és nemzetközi konferenciákon való publikálása. (Forrás: Finn-OTKA)
- A Magyar generatív történeti szintaxis 2. című projekt keretein belül elkészül a Káldi- és a Pesti Gábor-féle bibliafordítások betűhű változata. A középmagyar bibliafordítások normalizálását igyekszünk minél inkább megtámogatni automatikus nyelvtechnológiai módszerekkel. Ezáltal a Károli- és a Sylvester-féle bibliafordítások normalizálásával előreláthatóan végzünk az év során. Mivel 2017 a reformáció éve, kiemelt figyelemmel kezeljük a Károli Bibliát. Folyamatosan fejlesztjük a korpuszlekérdező felületeket és frissítjük a projekt weboldalát a frissen elkészülő szövegváltozatokkal. Digitalizált szövegelemek párhuzamos olvasását lehetővé tévő (és ezáltal a diakrón elemzést elősegítő) alkalmazást fejlesztünk. Terveink között szerepel a Universal Dependencies nemzetközi sztenderdhez való hivatalos csatlakozás, amelynek keretén belül szintaktikai annotációval kell ellátnunk a korpusznak legalább egy részét. (Forrás: MGTSZ2)
- Az uráli projekt keretein belül felépítjük a pilot projektben vállalt adatbázist minden nyelv minden nyelvallapotára (4000 token/kor/nyelv). Az egységes formátumra hozott fájlokat egyrészt szabadon elérhetővé tesszük a projekt honlapján, másrészt az általunk létrehozott összes adatot szeretnénk elérhetővé tenni egy nemzetközi nyelvi archívumon keresztül is, mint amilyen a The Language Archive által működtetett DOBES (Documentation of Endangered Languages) korpusz. Amennyiben az ugyanilyen témában beadott ERC pályázatot megnyerjük, a munkálatok nagyobb léptékben folytatódnak, a korpusz mérete pedig nagyságrendekkel bővülni fog. (Forrás: Urali)
- Terv a nemzetközi kapcsolatok továbbépítése a történeti kriptológia témájában. Amennyiben nyer a beadott COST projekt, annak keretei ezt lehetővé teszik. Ha nem, terveink között szerepel egy H2020-es pályázat beadása ebben a témában. Addig is legalább hazai szinten el lehet kezdeni a titkosításokból, a megfejtésükhöz használt kulcsokból, valamint a magyar nyelv különböző korokból származó szövegeiből való adatbázis felépítését. (Forrás: központi illetve a beadott COST projekt)
- Nyelvtechnológiai kutatások mélytanulási technikák alkalmazásával. Folytatódik a szóembeddingek nyelvészeti tartalmának, szemantikus hálókkal való kapcsolatának kutatása. Terv egy neurális hálóra épülő karakternormalizó fejlesztése. Terv a nem felügyelt gépi tanulási módszerekkel történő szemantikai modellezés kutatása, és az erre épülő alkalmazások fejlesztése. A kutatások előkészítésére a csoport néhány munkatársa mélytanulási továbbképzésen vett részt 2016-ban. (Forrás: központi)
- 2017-ben három jelentős budapesten megrendezett nemzetközi konferencia szervezésében vesz részt az osztály. Februárban az European Network of e-Lexicography zárókonferenciájára, majd a META-FORUM konferenciájára, végül szeptemberben a CLARIN éves konferenciájára kerül sor.
- Élénkülő együttműködés az MTA TTK Kognitív Idegtudományi és Pszichológiai Intézetével.
- A csoport kezelésében lévő korpuszok fejlesztése, ezzel kapcsolatos oktatás,

- ismeretterjesztés. (Forrás: központi)
- Az igei szerkezetek algebrai modelljének kutatása, publikáció megjelentetése. (Forrás: központi)
  - Az e-magyar rendszer népszerűsítése, a felhasználók körének bővítése, a rendszer nemzetközi publikálása. Együttműködés a NAVA-val a rendszer használatba vétele érdekében. (Forrás: központi)
  - A 2016-ban indult GINOP időarányos teljesítése. (Forrás: GINOP)
  - A Helyesírás tanácsadó portál üzemeltetése, fejlesztése. (Forrás: központi)
  - További pályázat, hazai és EU-s (H2020) pályázatok beadása.
  - A tudományos eredmények publikálása hazai és nemzetközi fórumokon.
  - Számítógépes nyelvészeti támogatás nyújtása nyelvészeti kutatásokhoz, nyelvi adatok szolgáltatása, kutatói kérések teljesítése a Magyar Nemzeti Szövegtár és az e-magyar rendszer tekintetében.

### **Matematikai Nyelvészeti Kutatócsoport**

A Matematikai Nyelvészeti Kutatócsoport 2016. december 31-ével megszűnt.

### **Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport**

- A nyelvi tanácsadó szolgálat működtetése és a nyelvi anyag tárolása, feldolgozása.
- Előadások és publikációk – a nyelvhasználat változása, új nyelvi jelenségek, kodifikációs problémák az írásgyakorlatban, az AkH. 12. változásainak érvényesítése.
- Nyelvi ismeretterjesztés – szakfolyóiratban és a médiában.
- Egyéni stilisztikai kutatások folytatása.
- A meglévő Utónévkereső adatbázisának frissítése, bővítése.
- Kapcsolat és megbeszélés a nemzetiségi önkormányzatokkal a névadás és a kérvények engedélyezésének alakulásáról.
- A névadással kapcsolatos tájékoztatás új formáinak megteremtése (honlap, ismeretterjesztő előadások). Az Utónévportál bővítése.